Adsorption of Proteins onto Charged Surfaces: A Monte Carlo Approach with Explicit Ions

A. H. JUFFER* **and P. ARGOS†**
EMBL Heidelberg, Postfach 10.2209, Meyerhofstrasse 1, D-69012 Heidelberg, Germany

J. DE VLIEG
Unilever Research Laboratory Vlaardingen, Postbus 114, NL-3130 AC Vlaardingen, The Netherlands

Received 24 May 1995; accepted 17 January 1996

ABSTRACT

A computer model has been developed to simulate the adsorption of proteins onto charged surfaces displaying an electric double layer. Coadsorption of ions onto the surface is included by means of explicit ions. Only electrostatic interactions are considered. Monte Carlo simulations in the canonical ensemble of the enzyme cutinase and 15 variants (modeled from the X-ray tertiary structure of the wild-type) were performed. Adsorption free energies for all variants were calculated by the thermodynamic integration method. Distributions of the electric moment and the vector pointing toward the protein active site and parallel to its central β-sheet were determined to elucidate the mean orientation of the protein with respect to the surface as a function of its distance from the surface. It was found that the free energy of adsorption varied linearly with the total charge of the protein, while the electric moment (dipole moment) had a second-order but significant effect. Though an increase of the electric moment generally resulted in a slightly increased affinity of the protein for the surface, close to the surface the mean force acting on the protein clearly varied linearly with the strength of the electric moment, such that a clear correlation between the latter and the protein orientation with respect to the surface could be established. Wild-type cutinase displayed the highest affinity for the charged surface amongst all proteins having the same total charge, even though it did not have the largest electric moment.

* Present address: The University of Calgary, Faculty of Science, Department of Biological Science, 2500 University Drive NW, Calgary, Alberta T2N 1N4 Canada.
† Author to whom all correspondence should be addressed.
E-mail: argos@EMBL-Heidelberg.de

Journal of Computational Chemistry, Vol. 17, No. 16, 1783–1803 (1996)

CCC 0192-8651 / 96 / 161783-21

Introduction

The study of adsorption of macromolecules onto surfaces is of major importance in the understanding of many biophysical processes. Several enzymes act only when bound to certain types of surfaces such as *lipases* at oil–water interfaces.[1] The association of particular proteins with a cell membrane triggers essential intracellular molecular processes.[2] Deposition and adhesion of oral microorganisms is the first step in the development of dental cavities.[3] Furthermore, many biotechnical processes depend on molecular interactions with a surface; for instance, ion-exchange chromatograpy.[2] In the pharmaceutical and food industries, proteins are being utilized as stabilizers for foams and emulsions.[4] It is clear that the design of biomaterials can be greatly enhanced by the comprehension of the physics involved in protein adsorption.

The complexity of these interfacial processes prevents a clear understanding of the macromolecule/surface interaction. Computer simulations in atomic detail are likely to reveal which types of intermolecular forces are of principal significance in determining the physical properties of a specific macromolecule/surface system. Experimental work carried out by Lohman et al.[4a] and Norde[4b] suggest that three major subprocesses contribute to the overall process of adsorption of proteins onto polymer surfaces (see also the reviewed by MacRitchie[5a] as well as by Norde and Lyklema[5b]): (1) dehydration effects; (2) redistribution of charged groups; and (3) rearrangements in the protein structure. Dehydration is closely related to the hydrophobic effect where the affinity between protein and surface is enhanced by hydrophobic contacts allowing a gain in entropy due to the release or displacement of water molecules from hydrophobic surfaces. Because a protein is predominantly stabilized by internal hydrophobic contacts, its structure may change under the influence of the surface leading to local rearrangements or even unfolding (sometimes called "surface denaturation"[4]). The greater freedom of the main chain results in entropy gain and thus the affinity of the protein for the surface is increased. However, the unfolding may have negative effects on the activity of an enzyme, especially when bound to the surface. If the unfolding is the major driving force for the binding at the surface, it can overcome repulsive electrostatic interactions (negative charge protein and a negative charged surface) between protein and surface and the binding of the protein to the surface is strong resulting in an almost negligible disassociation. The loss of the protein's translational and rotational freedom upon surface binding will counteract such hydrophobic effects. Finally, electrostatic interactions are of importance for adsorption phenomena. The binding of peripheral proteins to membranes is thought to be electrostatic in nature, because their isoelectrics points (pI) are generally different from 7.[6] A protein with a given overall total charge is attracted to a surface with opposite charge. Ions may be inserted into the layer between the protein and the surface to screen unfavorable electrostatic interactions. Such electrostatic effects largely depend on ionic strength and pH of the environment and may be strengthened by the dehydration effect.[6] The first was demonstrated by the experimental fact that the strength of the cytochrome c binding to charged lipid dispersions decreases with increasing ionic strength.[7] Close to the surface, the high concentrations of ions may lead to destabilization of the folded state of the protein ("salting out")[8] leading to increased affinity of the protein for the surface. In addition to the three subprocesses just mentioned, at high coverage of the surface, lateral interactions between adsorbed proteins will influence the amount of adsorbed proteins.[7] The so-called "soft" proteins (proteins which unfold at the surface at values of the pH different from their pI) will display a decreased plateau adsorption (the maximal amount of adsorbed proteins at a certain pH), because they take more space on the surface.[9]

Segregation of ions close to a charged surface leads to an electric double layer.[10] In this work, a computer model able to simulate a macromolecule in an electric double layer will be presented and will form the basis for a detailed investigation of the thermodynamic, structural, and dynamical properties of the macromolecule/surface system. Electric double layers have been studied in the past through continuum models based on the *nonlinear* Poisson–Boltzmann equation since ions are the most important component determining the properties of a double layer (Gouy–Chapmann theory).[10,11] In these models, the surface carries a uniform charge density and has no atomic detail. For 1:1 electrolytes, the electrostatic poten-

tial profile has been shown to follow a near exponential decay, whereas for 2:1 electrolytes the corresponding equation is much more complex. For low surface potentials (< 25 mV), the profile reduced to the Debye–Hückel approximation, $\psi_x = \psi_0 \exp(-\kappa x)$, where ψ_x is the potential at distance x from the surface, ψ_0 is the surface potential, and κ is the inverse Debye length. The continuum approach allows the calculation of the ionic distribution as a function of the distance to the surface, where there is a depletion of chargelike (with respect to the surface) ions at the surface, while oppositely charged ions tend to build up density close to the surface. Recently, Forsten et al.[12] used the finite-difference method to solve the nonlinear Poisson–Boltzmann equation for a membrane–electrolyte system in which the surface (membrane) was described as a regular array of charges.

In the presence of a macromolecule, the situation is yet more complex, because further interactions amongst the ions and the macromolecule and surface must be considered. Also, it may be necessary to include polarization effects from the presence of a dielectric boundary between the macromolecule and the solvent (and, possibly, the dielectric boundary between the charged surface and the solvent). It is generally not possible to solve analytically the continuum differential equations. For the *linear* Poisson–Boltzman equation, the potential problem can be reformulated into a set of linear integral equations at the surface of the macromolecule, numerically solvable with the boundary element method as shown by Yoon and Lenhoff.[13] These investigators also performed test calculations with ribonuclease A in 0.1 *M* ionic strength and a surface charge density of -0.018 C/m^2 resulting in interaction free energies of the order of 1 kT. Increasing the surface charge density to -0.1 C/m^2, a more realistic value, yielded energies up to 8 kT. However, with the linear Poisson–Boltzmann equation, the surface potential (ψ_0) is limited to about 25 mV, which, for a 0.1 *M* NaCl solution at 298 K, corresponds to a surface charge density of about 0.02 C/m^2 (the dielectric constant of water was taken to 78.5). Roth and Lenhoff[14] extended the continuum approach of Yoon and Lenhoff[13] by including van der Waals interactions between the surface and macromolecule through the Hamaker equation applicable to spheres. However, such dispersion interactions are thought not to be important for adsorption.[5] Since the calculations were computationally demanding, instead of taking the full protein charge distribution, Roth and Lenhoff[14] placed a monopole and occasionally a dipole moment at the center of the protein which was assumed to be a sphere. Roush et al.[15a] considered the interaction of the soluble tryptic core of rat cytochrome b_5 and a simulated anion-exchange surface through application of the linearized finite-difference Poisson–Boltzmann equation. The ionic strength was set to 0.1 *M* and atomic detail for the surface, carrying a charge density of 0.1 C m^{-2}, was included. They found a preferred orientation with respect to the surface of the protein, although they were not able to consider all possible orientations of the protein because of the computational demands. The determined preferred orientation was, however, in correspondence with experimental results.[15b]

One may use a formulation in which the solvent is explicitly taken into account. Such a model is more realistic and flexible than one based purely on a full continuum description of the solvent. For explicit solvent, there are no restrictions on the ionic strength and surface charge densities. Correlations between the system components (ions, water, surface, macromolecule, etc.) can be properly considered which is not the case in the continuum description of the solvent. These correlations become more important at higher ionic strength (> 0.01 *M*),[16,17] although for the binding of the radical O_2^- to the enzyme superoxide dismutase it was noted that, upon comparison with Monte Carlo results, the assumptions inherent in the Poisson–Boltzmann were acceptable for that particular application even at an ionic strength as high as 0.5 *M*.[18,19] Various types of ions (e.g., mixtures of 1 +, 2 +, 1-charged ions) or other types of solvent molecules (e.g., small organic molecules, surfactants) can also be included. Furthermore, for adsorption, the complicated nature of the interaction of the two double layers can be treated in a consistent manner. The coadsorption or binding of ions in the diffuse ionic atmosphere around the protein and next to the charged surface, which have impact on the adsorption of a particular protein,[4,5,20] can thus be properly included.

The adsorption model presented here is based on atomic detail for ions and protein and a continuum description for water, whose treatment at the atomic level is desirable to account fully for the hydrophobic effect,[4,5] although the computational cost is prohibitive, and the surface is described as a uniform surface charge density. Such an approach was also employed by Jarayam et al.[21] to study the effect of ions in protein–DNA interac-

tions; electrostatic interactions were screened by means of an effective distance-dependent screening function to model the effect of water and the existence of a dielectric boundary between solvent and solute. In this work, a single dielectric constant (of water) is applied to the system, such that no polarization can be considered, even though Jayaram et al.[21c] found that different dielectric models for the solvent may effect, for instance, the counterion distribution around macromolecules. To our knowledge, computer simulation of the adsorption of proteins onto surfaces at this level of detail for the solvent has not yet been performed, although its importance has been recognized.[15]

A major goal of the present model is the calculation of the adsorption free energy, the only thermodynamic quantity which is able to predict if a process (here the physical adsorption of a protein) will occur. It is important to discriminate among several variants of a protein so that is possible to consider the effect of a change in the charge distribution of a protein molecule on its adsorption behavior. Because the model allows prediction of adsorption ability for various residue substitutions, it should prove its significance in protein engineering and design. The enzyme cutinase[22] and 15 variants were considered as test cases.

CUTINASE

The enzyme cutinase[22] (see Fig. 1 for its structure) displays properties of both an esterase and a lipase. A lipase attacks triacylglycerols and acts only at an oil–water interface mainly by hydrolysis of ester linkages.[1] The interfacial activation arises from a hydrophobic lid which normally covers the active site but is displaced on binding to a substrate. Esterases do not depend on an interface but do display lipase activity. Cutinase relies on an interface but without interfacial activation. Its active site, containing the catalytic residues Ser-120, His-188, and Asp-175, is solvent accessible and not covered by a lid. On association with a lipid surface, small displacements of hydrophobic loops close to the active site help form the binding site, albeit soluble substrates are acceptable also.

Wild-type cutinase, whose tertiary structural data is contained in entry "1CUS" of the Brookhaven Protein Data Bank,[23] displays a very asymmetric charge distribution such that the electric potential around the protein resembles that of a dipole. It would be expected that long-range electrostatic interactions (especially the dipole component) are of major importance in steering the molecule (active site) into the most effective orientation with respect to a charged surface. Cutinase thus becomes an excellent test molecule in developing computer simulation tools to study adsorption phenomena in detail.

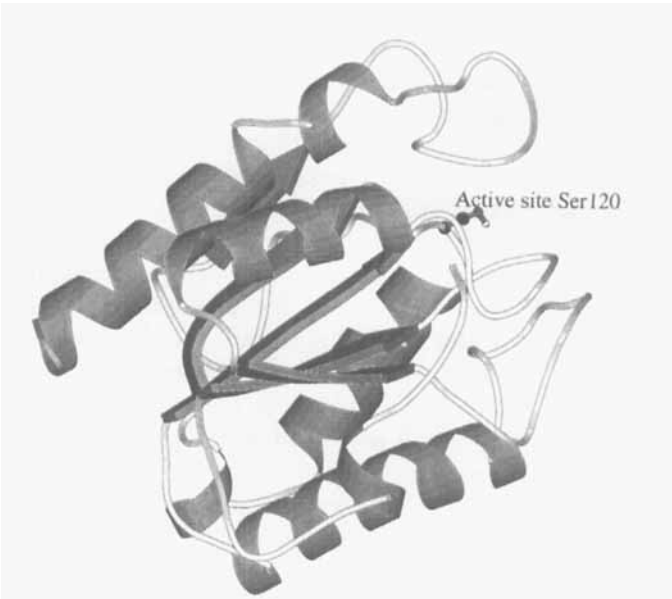

FIGURE 1. Schematic representation[53] of the backbone of cutinase. The molecule has five almost parallel β-strands of which the central one is pointing toward the crevice in which the active Ser-120 is located.

Methods

The model employed in this work is based on the work of Torrie and Valleau,[24] who investigated the properties of an electric double layer using a Monte Carlo approach, but did not consider a macromolecular application. The model used here is illustrated in Figure 2; details will be introduced throughout the ensuing text.

A box with dimensions $L \times L \times W$ (L is the length in the x- and y- directions and W is the width in the z-direction) containing a rigid macromolecule and explicit ions (in this work only Na^+ and Cl^-) is placed next to a uniformly charged surface (indicated with sub/superscript B) carrying a (negative) surface charge density σ_B. The number of excess ions compensates for the charge density on the surface and the total charge of the macromolecule such that the overall total charge of the system is zero. The surface is taken to be parallel to the xy-plane and intersects the positive z-axis at $z = W/2$. The origin is taken at the box center. Table I defines various parameters, their

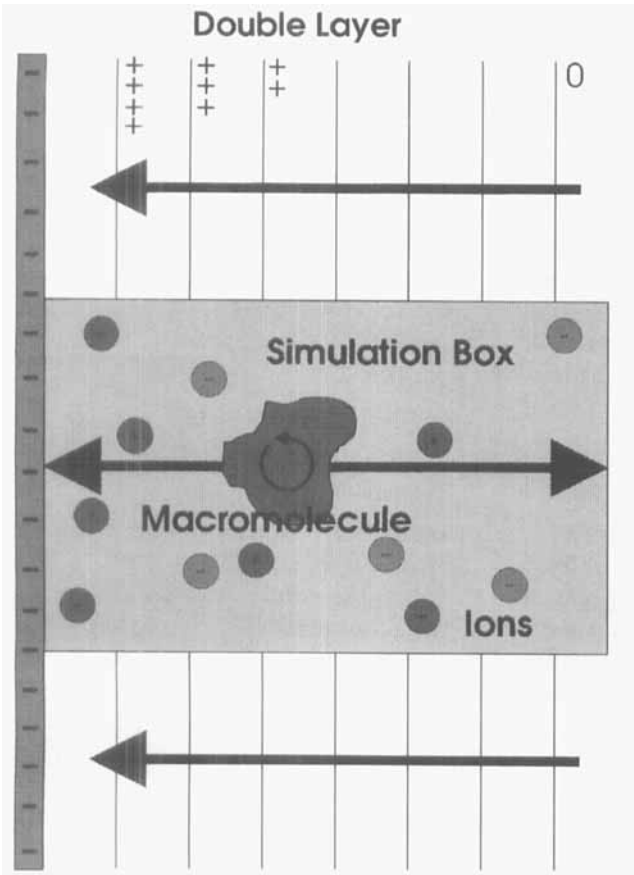

FIGURE 2. Simulation model for adsorption (see the text for a description).

symbolic notation, and explicit values used in this work.

ADSORPTION FREE ENERGY

To calculate the free energy of adsorption, the classical partition function Q in the canonical ensemble[25] is:

$$Q = \frac{1}{\Pi_s \Lambda_s^{3N_s} N_s!} \int \exp\left(-\frac{U_N}{kT}\right) d\{\mathbf{N}\} \qquad (1)$$

Here, {**N**} refers to set of generalized coordinates for N molecules. U_N is the total potential energy of the system, including the interaction with a surface carrying a uniform surface charge density, T is the temperature and k is Boltzmann's constant. Λ_s is a constant with a similar meaning as the de Broglie constant and arises from an integration over generalized impulses, the subscript s refers to an atom or molecular species (here, Na^+, Cl^- and protein where water is considered as a background continuum in which the ions and protein are immersed), and N_s is the number of molecules or atoms of type s.

TABLE I.
Overview of all Simulation Parameters.

Meaning	Symbol	Value
Dielectric constant	ε	78.5
Width of Stern layer	δ	0.2125 nm
Temperature	T	298 K
Surface charge density	σ_B	-0.2661 C m^{-2}
		-1.667 e nm^{-2}
Distance of closest approach	σ	0.425 nm
Width simulation box	W	36.0 nm
Length simulation box	L	18.0 nm
Number of virtual planes	M	72
Total number of charges	N	4810[a]
Total number of ions	n_{ions}	1943[a]
Ionic strength	I	0.1 M
Bulk number density	ρ	0.06 nm^{-3}

Most parameters were taken from Torrie and Valleau,[24] but they did not specify the number of virtual planes used. The value of 72 results in a distance of about 0.5 nm (about one ion layer) between the planes. The diameter of the proteins is a bout 5.5 nm and the range of the ion–ion correlation for 0.1 M NaCl is 3–4 nm.[36] A box length of 18 nm results in about 6 nm of space between the protein and the edge of the box in the x- and y-direction. For a simulation of a nonuniform system with long-range electrostatic interaction with periodic boundary conditions together with the minimum image approximation, it is preferable to use a large box.[35,37] Because *a priori* it is not known at what distance from the surface the protein will still experience a force, a rather large value of the box width (z-direction) was taken. This also ensures that errors resulting from missing interactions between the ions beyond $z = -1/2W$ and all ions and the protein inside the box are minimized. Such large box dimensions were not used to Torrie and Valeau.[24] Realistic surface charge densities[14] start at a value of -0.1 C m^{-2}. Therefore, the value of -0.2661 C m^{-2} is a reasonable choice and was also used by Torrie and Valeau[24] (among other values).
[a]For wild-type cutinase.

The protein is treated as rigid such that a total of six generalized coordinates are required to specify its position and its orientation[26]; three generalized coordinates correspond to the cartesian coordinates of its center of mass **R**; and three others correspond to the Euler angles φ, θ, and ψ. The partition function can now be written as:

$$Q = \frac{1}{\Pi_s \Lambda_s^{3N_s} N_s!} \int d\{\mathbf{r}\} \int d\mathbf{R} \int d\varphi\, d\theta\, d\psi \times \sin\theta \exp\left(-\frac{U(\{\mathbf{r}\}, \mathbf{R}, \varphi, \theta, \psi)}{kT}\right) \qquad (2)$$

where {r} corresponds to the Cartesian coordinates of ions. Any translations of the protein parallel to the surface are assumed not to change the potential energy such that Q becomes:

$$Q = \frac{L^2}{\Pi_s \Lambda_s^{3N_s} N_s!} \int d\{\mathbf{r}\} \int dz \int d\varphi\, d\theta\, d\psi \times \sin\theta \exp\left(-\frac{U(\{\mathbf{r}\}, z, \varphi, \theta, \psi)}{kT}\right) = \frac{L^2}{\Lambda_x^3} \int dz\, Q'(z) \quad (3)$$

where:

$$Q'(z) = \frac{8\pi}{\Lambda_\theta \Pi_i \Lambda_i^{3N_i} N_i!} \int d\{\mathbf{r}\} \int d\varphi\, d\theta\, d\psi \times \sin\theta \exp\left(-\frac{U(\{\mathbf{r}\}, z, \varphi, \theta, \psi)}{kT}\right) \quad (4)$$

The symbol Λ_x in eq. (3) is the de Broglie constant and arises from the kinetic part of the translational degrees of freedom of the protein (as indicated by the subscript x). The index i in eq. (4) refers to ions only. The symbol Λ_θ in eq. (4) is a constant with a similar meaning as the de Broglie constant, but arises from the kinetic part of the orientational degrees of freedom of the protein (as indicated by the subscript θ, not to be confused by the Euler angle θ) (see Hill[25] for more details). The exact form of these constants is not essential for what follows.

The function $Q'(z)$ is a partition function of the system when the macromolecule is located at a certain point along the z-axis and includes all possible orientations of the protein and all possible positions of the ions (Fig. 2; the z-axis corresponds to the arrow through the center of the box). It is related to a free energy $A'(z)$ according to:

$$A'(z) = -kT \ln Q'(z) \quad (5)$$

From eqs. (4) and (5), one derives:

$$\frac{\partial Q'(z)}{\partial z} = -\frac{8\pi}{kT\Lambda_\theta^3 \prod_i \Lambda_i^{3N_i} N_i!} \int d\{\mathbf{r}\} \int d\varphi\, d\theta\, d\psi \times \sin\theta \exp\left(-\frac{U(\{\mathbf{r}\}, z, \varphi, \theta, \psi)}{kT}\right) \frac{\partial U}{\partial z} \quad (6)$$

and:

$$\frac{\partial A'(z)}{\partial z} = -kT \frac{1}{Q'(z)} \frac{\partial Q'(z)}{\partial z} = \left\langle \frac{\partial U(z)}{\partial z} \right\rangle = -F'(z) \quad (7)$$

$F'(z)$ is a mean force canonically averaged over all possible configurations of all the ions **and** over all possible *orientations* of the macromolecule. Therefore, $A'(z)$ may be considered as a potential of mean force.[25]

Now, the quantity $\Delta A'$:

$$\Delta A' = \int_{-\infty}^{z_0} dz \frac{\partial A'(z)}{\partial z} = \int_{-\infty}^{z_0} dz \left\langle \frac{\partial U(z)}{\partial z} \right\rangle \quad (8)$$

is the change in the total free energy in moving the macromolecule from bulk solution at $z = -\infty$ to a point $z = z_0$. Since the origin was placed at the center of the box, the quantity $\langle \cdots \rangle$ in eq. (8) is negative for an attractive interaction between the protein and the charged surface. If $z = z_0$ corresponds to a point close to the surface, $\Delta A'$ may be related to the free energy of adsorption. No assumptions have been made regarding the total potential energy of the system such that expression (8) is exact.

Eq. (8) exemplifies so-called "thermodynamic integration."[27] The parameter z (for which the symbol λ is commonly used) refers to a path of integration. Any path can be taken since the free energy is a state function. The most striking feature of the last equation involves the calculation of an ensemble average, a much easier calculation than that required for the partition function.

If:

$$U(\{\mathbf{r}\}, z, \varphi, \theta, \psi) = U_0(\{\mathbf{r}\}) + U'(\{\mathbf{r}\}, z, \varphi, \theta, \psi) \quad (9)$$

where U_0 is the potential energy of the ions only (*excluding the interaction with the macromolecule*) and U' is the *interaction* potential energy of the macromolecule with the rest of the system (*including* the interaction between the ions and the macromolecule), then:

$$\frac{\partial U}{\partial z} = \frac{\partial U'}{\partial z} \quad (10)$$

INTERACTIONS

The interaction potential $u_i^B(d_i^B)$ ($d_i^B = W/2 - z_i$, $-W/2 < z_i < W/2$ and corresponds to the dis-

tance of ion i from the surface) of a ion with the charged surface is,[28] in MKS units:

$$u_i^B(d_i^B) = -q_i\frac{\sigma_b}{2\varepsilon\varepsilon_0}d_i^B + C, \qquad \text{if } d_i^B > \delta$$
$$= 0 \qquad \text{if } d_i^B < \delta \quad (11)$$

Here, q_i is the charge of the ion, ε_0 is the vacuum permittivity ($8.854187817 \times 10^{-12}$ F/m), and ε is the dielectric constant of water. The constant C depends on the choice of the reference point where the electrostatic potential of the surface is zero [taken here at $d_i^B = 0$ implying $C = 0$, although the exact choice is not critical, because, in the Monte Carlo procedure employed, only differences in $u_i^B(d_i^B)$ are considered]. The parameter, δ, is the width of a Stern layer[10,11] and corresponds to the distance of closest approach between an ion within the box and the charged boundary. Eq. (11) also applies to the macromolecule which is considered as a set of partial charges at atomic sites. The value of δ is taken to be the same for all interactions of the type considered in eq. (11).

The interaction $u_{ij}^{PM}(r_{ij})$ between an ion pair (i, j) separated by a distance r_{ij} according to the primitive model (PM) for electrolyte solutions can be expressed as[16,17]:

$$u_{ij}^{PM}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon\varepsilon_0 r_{ij}} \qquad \text{if } r_{ij} > \sigma,$$
$$= 0 \qquad \text{if } r_{ij} < \sigma \quad (12)$$

where q_i represents the charge of ion i and σ is the distance of closest approach between two ions. For interactions between the macromolecule and ions (also including those that possibly fall within the protein), the same equation is used such that, if i represents an ion, then the index j runs over all partial charges of the macromolecule. The same value for σ is taken for all interactions of the type considered in eq. (12).

Polarization effects due to the difference in permittivity between solvent and macromolecule and as well as the surface and solvent are not included such that in this work a single dielectric constant (of water) is applied to the full system. It is often assumed that the dielectric constant of a protein is about 2 to 4,[29,30] which yields credible results such as in the calculation of the acid dissociation constants.[31] It has been shown that electrostatic potentials based on continuum electrostatics depend critically on the exact location of the dielectric boundary; its effect is magnified close to the boundary.[21a,30] Several other studies have utilized only a single dielectric constant (i.e., neglecting polarization effects). For instance, Soumpasis[32] was able to predict correctly the B → Z transition of DNA using a potential of mean force approach, while Jönsson and Svensson[33] accurately computed equilibrium constants for the binding of ions (Ca^{2+}) to proteins where the solvent contained explicit ions but water was considered as a continuum. Recently, using molecular dynamics, the dielectric constant of proteins was estimated to be around 30[34] instead of 2 to 4.[29a] In this work, a single dielectric constant is applied mainly for reasons of computational efficiency although it is expected that polarization effects do play a role in adsorption phenomena.

In a "pure" double layer (without a macromolecule), the equilibrium distributions of positive and negative ions are asymmetric. An electric moment exists which points toward the charged boundary (two large arrows outside the simulation box in Fig. 2).[10,11,24] Long-range electrostatic interactions require consideration of the interactions between ions within the box and ions outside the box. The concept of "virtual planes" (indicated by the parallel lines in Fig. 2) was introduced by Torrie and Valleau[24] to mimic the asymmetric charge distribution outside the box. These planes, which are positioned parallel to the charged surface, but do not overlap the central box (they have "holes" of dimension L by L), are charged (as indicated by the "+" symbols in Fig. 2) and opposite in sign with respect to the charged surface. The surface charge density of a virtual plane is smaller if the plane is located further from the surface. These surface charge densities are sampled during the simulation of a pure double layer and they correspond to the average ion density in the space between two successive virtual planes. The charge densities of the planes are taken to be static (although they can fluctuate) during the simulation when the macromolecule is present in the double layer.

In Appendix A, it is shown that the interaction of a charge i (ion or [partial] charge of the macromolecule) with a virtual plane P_j is:

$$u_i^{P_j}(d_i^{P_j}) = -q_i\frac{2\sigma_{P_j}}{\pi\varepsilon\varepsilon_0} \times \int_0^{1/4\pi}\left(\frac{\frac{1}{4}L^2 + (d_i^{P_j})^2\cos^2\varphi}{\cos^2\varphi}\right)^{1/2} d\varphi + C \quad (13)$$

Here, σ_{P_j} is the surface charge density of the plane, $d_i^{P_j} = |z_i - z^{P_j}|$ is the distance between the charge and the plane P_j, and L is the length of a side in a square corresponding to that part of the virtual plane which would be inside the simulation box if the virtual plane would overlap the box. The constant C in eq. (13) has a similar meaning as in eq. (11) (see Appendix A). The integral can be solved with a numerical quadrature.

The total potential energy $U_N(\{r\}, z, \varphi, \theta, \psi)$ of the system now becomes:

$$U(\{\mathbf{r}\}, z, \varphi, \theta, \psi) = \sum_{i}^{N_{cg}} u_i^B(d_i^B) + \sum_{j}^{M} \sum_{i}^{N_{cg}} u_i^{P_j}(d_i^{P_j}) + \sum_{i<j}^{N_{cg}} u_{ij}^{PM}(\mathbf{r}_{ij}) \quad (14)$$

N_{cg} refers to all charges in the system and M is the total number of virtual planes.

If the expression for the total potential energy is compared with eq. (9), then

$$U'(\{\mathbf{r}\}, z, \varphi, \theta, \psi) = \sum_{l}^{n_{cg}} u_l^B(d_l^B) + \sum_{j}^{M} \sum_{l}^{n_{cg}} u_l^{P_j}(d_l^{P_j}) + \sum_{k}^{n_{ions}} \sum_{l}^{n_{cg}} u_{kl}^{PM}(r_{kl}) \quad (15)$$

The total number of ions in the system is n_{ions} and n_{cg} is the total number of partial charges of the macromolecule ($N_{cg} = n_{ions} + n_{cg}$). Both d and r_{kl} depend on z, φ, θ, and ψ.

In Appendix B, the expression for $\partial U'/\partial z$ is shown to be:

$$\frac{\partial U'}{\partial z} = \frac{\sigma_B}{2\varepsilon\varepsilon_0} Q + \sum_{j=1}^{m} \sum_{l=1}^{n_{cg}} q_l \frac{2\sigma_{P_j}}{\pi\varepsilon\varepsilon_0} \times \sin^{-1}\left(\frac{(d_l^{P_j})^2}{\frac{1}{4}L^2 + (d_l^{P_j})^2} \frac{1}{2}\sqrt{2} \right) - \sum_{k=1}^{n_{ions}} \sum_{l=1}^{n_{cg}} u_{kl}^{PM}(r_{kl}) \frac{z_k - z_l}{r_{kl}^2} \quad (16)$$

The total charge Q of the macromolecule is thus of major importance for the adsorption free energy. The second and third terms reflect the effect of the charge distribution of the macromolecule on adsorption. The second term cancels the first term if the macromolecule is far away from the surface (further away than the dimension of the hole) since the virtual planes act on the macromolecule as if they have no holes. The last term is the z-component of a force acting on the charges of the macromolecule as a result of the ions in the system. This term approaches zero if the distance between the macromolecule and the charged surface becomes infinite.

SIMULATION DETAILS

The system is simulated in the canonical ensemble with the Monte Carlo method[35] using the parameters given in Table I. To perform the thermodynamic integration in eq. (8), the macromolecule is placed at several distances d [d = 4, 6, 8, 12, 16, 20, and 24 nm corresponding, respectively, to z = 14, 12, 10, 6, 2, −2, and −6 in eq. (8)] with respect to the boundary. The value of $\langle \partial U'/\partial z \rangle$ at each distance is determined and the value for the adsorption free energy is calculated by means of a numerical integration procedure (see below).[27] The closest possible distance of the macromolecular center of mass to the surface is 4 nm. Because the radii of the cutinase variants are about 2.5 nm, there remain around 1.5 nm between the surface of the protein and the charged surface. Since the charged surface is currently described as a uniform surface charge density, atomic details for direct interactions (e.g., Coulomb interaction or Lennard–Jones interaction) at close distances between surface and protein cannot be considered. Therefore, the protein cannot not be placed too close to the surface.

For the ions, a trial state was generated by choosing the x (and similarly for the y) coordinate from a small interval with length $1/2L$ around the current x coordinate. The z coordinate (perpendicular to the charged boundary) was chosen uniformly from the interval $[1/2W - \delta, -1/2W + \delta]$. The orientations of the macromolecule are sampled through the Euler angles φ, ψ, and θ.[26,35] The cosine of θ, instead of θ, was sampled uniformly in the interval $[-1, 1]$.[35] The angles φ and ψ were sampled between $[-1/2\pi, 1/2\pi]$ and their periodicity was employed. If the protein was selected for a random orientational move, any ion within a distance range (the radius of the macromolecule) with respect to the center of mass of the protein was also rotated along with the macromolecule, necessitated by immediate overlap of the rotated macromolecule with nearby ions. Because no preferential sampling for the ions was observed following this procedure, we assumed that there was no need to modify the underlying transition ma-

trix. On average, an acceptance ratio of about 50% was observed with the simulation parameters in Table I.

All interactions within the central box were calculated in the minimum image convention.[35] Periodicity is imposed parallel to the charged surface. Any interaction of charges (ions or partial charge of the macromolecule) inside the box with ions beyond $z = -1/2W$ were neglected. No virtual planes were taken beyond $z = -1/2W$ simply because they would be uncharged (as indicated by "0" in Fig. 2). These approximations may result in errors in the properties of ions inside the box and close to $z = -1/2W$, since they do not interact with ions just outside the box. If the system is sufficiently large (much larger than the range of the ion–ion correlation, for 0.1 M NaCl about 3–4 nm[36]), it is not expected that these errors will influence the properties of the macromolecule providing it was not too close to $z = -1/2W$. In the present simulations (W = 36.0 nm), these problems are unlikely.

A Monte Carlo simulation was initiated by placing the protein at one of the distances mentioned earlier in an already equilibrated double layer of 0.1 M NaCl. The latter was generated by a $1.25 \times 10^{+6}$ step Monte Carlo simulation (the first 250,000 steps were considered as an equilibration period) from which the surface charge densities of the virtual planes were also determined. For the simulation of a protein in a double layer the first 500,000 steps were considered as an equilibration period. The total number of Monte Carlo steps varied between $2.5 \times 10^{+6}$ and $10.0 \times 10^{+6}$. For the cases in which the protein was placed close to the surface ($4 \leq d \leq 8$ nm), the number of steps was $5–7.5 \times 10^{+6}$ (in one case it was extended to $10 \times 10^{+6}$). For all other cases ($d > 8$ nm), only $2.5 \times 10^{+6}$ Monte Carlo steps were performed. Every 5826 (about threefold the number of molecules in the system) steps, the configuration was written to a trajectory used for later analysis.

CUTINASE VARIANTS

The simulations were performed for wild-type cutinase[22] and 15 variants. A set of 12 variants was designed with the same total charge as the wild-type (total charge is +1 at pH 7); however, they varied in the strength of their electric moments. A set of three variants was designed with different total charges ranging from −1 to +5. Residue substitutions of all variants relative to the wild-type are given in Table II, whereas Table III lists

TABLE II.
List of Residue Mutations for All Variants.

Protein	Mutation(s)
Wild-type	None
n172k	n172k
t45k	t45k_s92r_e201k
r17e	r17e
a1	s92r_199k_g100r_n58r_g67d_a62e_i24e_p162e
a2	s92r_s61r_n58r_a142k_g75d_a62e_s57d_i24e
a3	a142k_s61r_n58r_g100r_n172d_t80d_g75d_n161d
a5	199r_s61r_a142k_s92r_n161d_p163e_g75d_s57d
b1	s92r_199k_g67d_a62e
b2	s61r_n48r_a62e_g75d
b3	s61r_a142k_t80d_n172d
b5	199r_s61r_n161d_p163e
c1	r208g_k140q_r138q_k65a_d83s_d165s_e131s_e201n
c2	r196s_r208g_r156l_k140q_d134s_d139n_d21n_d66s
c3	r17n_r156l_r196s_r168q_d33s_d111n_d66s_e205a
c5	r78s_r88a_r156l_r168q_d33s_d66s_d83s_e131s

The first column gives the variant code, while the second column lists the residue mutations. All newly designed variants were locally energy minimized with InsightII.[38] For the a-, b-, and c-variants, the changes in the charge distribution of the wild-type were performed on the surface of the wild-type. These variants have the same total charge as the wild-type (see Table III). For the a-variants eight charges were added: four positive and four negative. For the b-variants there were four charges added: two positive and two negative. For the c-variants there were eight charges removed: four positive and four negative. The partial charges for all the atoms (which included all hydrogen atoms) or a variant were taken from InsightII[38] and correspond to the charge distribution at pH 7.

TABLE III.
Some Properties of Variants.

Protein	q (e)	μ (10^{-27} C m)	μ (e nm)	Q (10^{-37} C m^2)	Q (e nm^2)	Fraction hydrophilic area
wt[a]	1	1.46	9.05	48.1775	30.07	0.416
n172k	2	1.57	9.74	47.9852	29.95	0.416
t45k	5	1.61	9.98	55.3232	34.53	0.412
r17e	−1	1.37	8.5	33.838	21.12	0.417
a1	1	2.16	13.44	62.6932	39.13	0.433
a2	1	1.52	9.45	52.0547	32.49	0.433
a3	1	1.06	6.58	53.1923	33.2	0.429
a5	1	1.53	9.5	42.5699	26.57	0.431
b1	1	2.23	13.85	59.5369	37.16	0.42
b2	1	1.54	9.57	57.6624	35.99	0.421
b3	1	0.82	5.11	55.4353	34.6	0.422
b5	1	1.75	10.87	48.5139	30.28	0.424
c1	1	2.2	13.65	54.8906	34.26	0.401
c2	1	1.42	8.83	56.0281	34.97	0.397
c3	1	0.69	4.26	13.7307	8.57	0.399
c5	1	1.13	7	33.9662	21.2	0.395

The first column is the cutinase variant code, the second column lists the total charge q (in proton charge e), the third and fourth columns contain the strength of electric moment μ in e nm and 10^{-27} C m respectively, and the fifth and sixth column list the strength of quadrupole moment Q in 10^{-37} C m^2 and e nm^2, respectively. Precise definitions of the electric moment and the quadrupole moment are given in the text. The last column lists the fraction of hydrophilic area of each variant which was calculated from the solvent-accessible surface[42] and atomic solvation parameters (ASP).[43] If an ASP of a particular atom was negative, it was assigned to be a hydrophilic atom.

[a] For wild-type cutinase.

some of their properties. InsightII[38] was used for the design of the 15 variants; some details can be found in the footnote to Table II.

ERROR ANALYSIS

It is difficult to determine the error in a calculated free energy difference. Many different paths (different from the path perpendicular to the charged surface as used in this work) should be included and then the variations in the calculated free energies considered. The computer demands are prohibitive. However, it is possible to calculate the error in the equilibrium values of $\langle \partial U'/\partial z \rangle$ (see Allen and Tildesley[35] and references therein). Because the values of $\langle \partial U'/\partial z \rangle$ are determined from a possibly correlated data series, the error in each mean is not simply given by $\sigma^2(\langle A \rangle_{run}) = \sigma^2(A)/N$, where A is the quantity of interest (here $\langle \partial U'/\partial z \rangle$), $\sigma^2(A)$ is the regular variance of the data series, and N is the length of the data series, implying all measurements of A are statistically independent. Instead, the sequence is divided into blocks or subseries of equal length (if the last block had a length different from the others, it was not considered). From each block, an average $\langle A \rangle_b$ is calculated. The mean value for all the blocks is then used to determine the variance $\sigma^2(\langle A \rangle_b)$ of the mean of the full data series according to:

$$\sigma^2(\langle A \rangle_b) = \frac{1}{n_b} \sum_{b=1}^{n_b} (\langle A \rangle_b - \langle A \rangle_{run})^2 \quad (17)$$

Here, n_b is the number of blocks. The quantity $\sigma^2(\langle A \rangle_b)$ is expected to be inversely proportional to the block length l_b for large blocks. In this limit, the means for each block are statistically independent. The so-called statistical inefficiency, s, defined as:

$$s = \lim_{l_b \to \infty} \frac{l_b \sigma^2(\langle A \rangle_b)}{\sigma^2(A)} \quad (18)$$

becomes the minimal block length such that the block averages are statistically independent. Given s, the error $\sigma(\langle \partial U'/\partial z \rangle)$ in the mean of $\partial U'/\partial z$ becomes:

$$\sigma\left(\left\langle \frac{\partial U'}{\partial z} \right\rangle\right) = \left(\frac{s}{N}\right)^{1/2} \left\langle \frac{\partial U'}{\partial z} \right\rangle. \quad (19)$$

Results

PURE DOUBLE LAYER

To simulate a macromolecule in an electric double layer, the "pure" double layer (without the macromolecule) must be simulated first. Figure 3 shows the density profile versus the distance from the surface. Calculations were performed for 0.1 *M* NaCl (bulk number density 0.06 nm^{-3}) and a surface charge density of -0.2661 C/m^2 (see Table I). Clearly, the positive ions have a tendency to accumulate near the negatively charged surface where, correspondingly, the negative ions are depleted. Bulk density is reached at about 5 nm from the surface. These results are in agreement with Torrie and Valeau.[24]

CUTINASE

Adsorption Free Energy

A priori, it is not known at which distances the protein experiences an interaction with the surface. Figure 4 shows, for all variants, the behavior of $\langle \partial U'/\partial z \rangle$ as a function of the distance d from the protein center of mass to the surface. Although $\langle \partial U'/\partial z \rangle$ does approach zero for larger d, at $d = 24$ nm the value of $\langle \partial U'/\partial z \rangle$ is not yet exactly zero, meaning that the proteins still experience an attractive force (except for variant r17e [Fig. 4A], which experiences a repulsive force) toward the surface at this large distance. The first three distances ($d = 4$, 6, and 8 nm) have been sampled quite accurately. Any fluctuations beyond this range might be artificial due to potentially insufficient sampling ($2.5 \times 10^{+6}$ Monte Carlo steps) necessitated by the prohibitive computer costs. Nonetheless, the most substantial segment of the free energy of adsorption is related to distances between 0 and 10 nm.

Because no values for $\langle \partial U'/\partial z \rangle$ were "measured" at larger distances ($d > 24$ nm), an approximation was used—namely, that according to Prony's method,[39] of the form:

$$f(x) = C_1 e^{a_1 x} + C_2 e^{a_2 x} + \ldots + C_n e^{a_n x} \quad (20)$$

which was employed to calculate the integral along the path z for d larger than 24 nm (tail value). The a_i values are determined from the roots (which may be complex) of a polynomial of order n, and the C_i values can be calculated by a least-squares fit using n exponential functions and exploiting all the points $(\langle \partial U'/\partial z \rangle_j, d_j)$. An extended trapezoidal rule[40] in combination with an interpolation scheme according to Neville's algorithm[40] to cal-

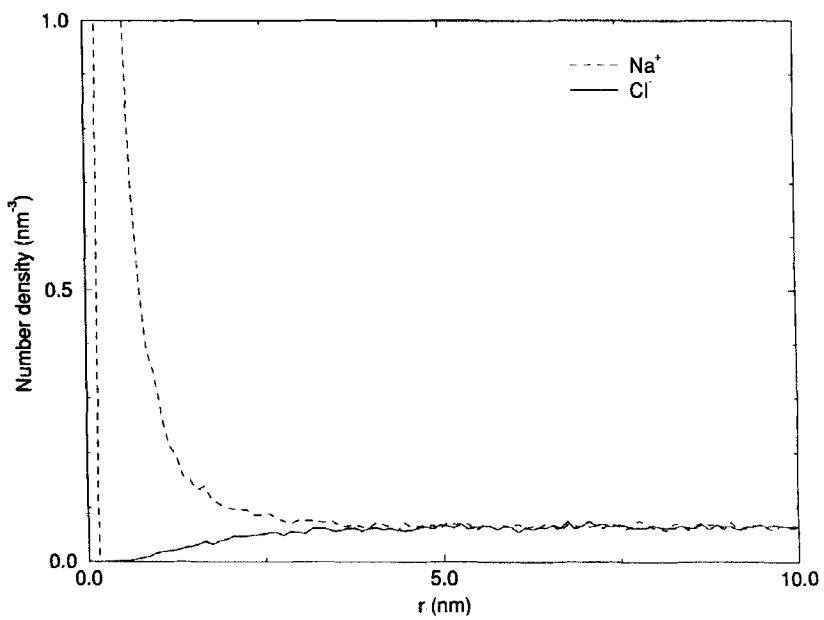

FIGURE 3. Calculated ion (Na^+ and Cl^-) number density, $\rho(d)$, in the double layer as a function of the distance d from the charged surface. The double layer was simulated for 0.1 *M* NaCl, and a surface charge density of -1.660 e nm^{-2} (-0.2661 Cm^{-2}), and with the parameters given in Table I. The number of Monte Carlo steps was $1.25 \times 10^{+6}$. The first 250,000 steps were used for equilibration.

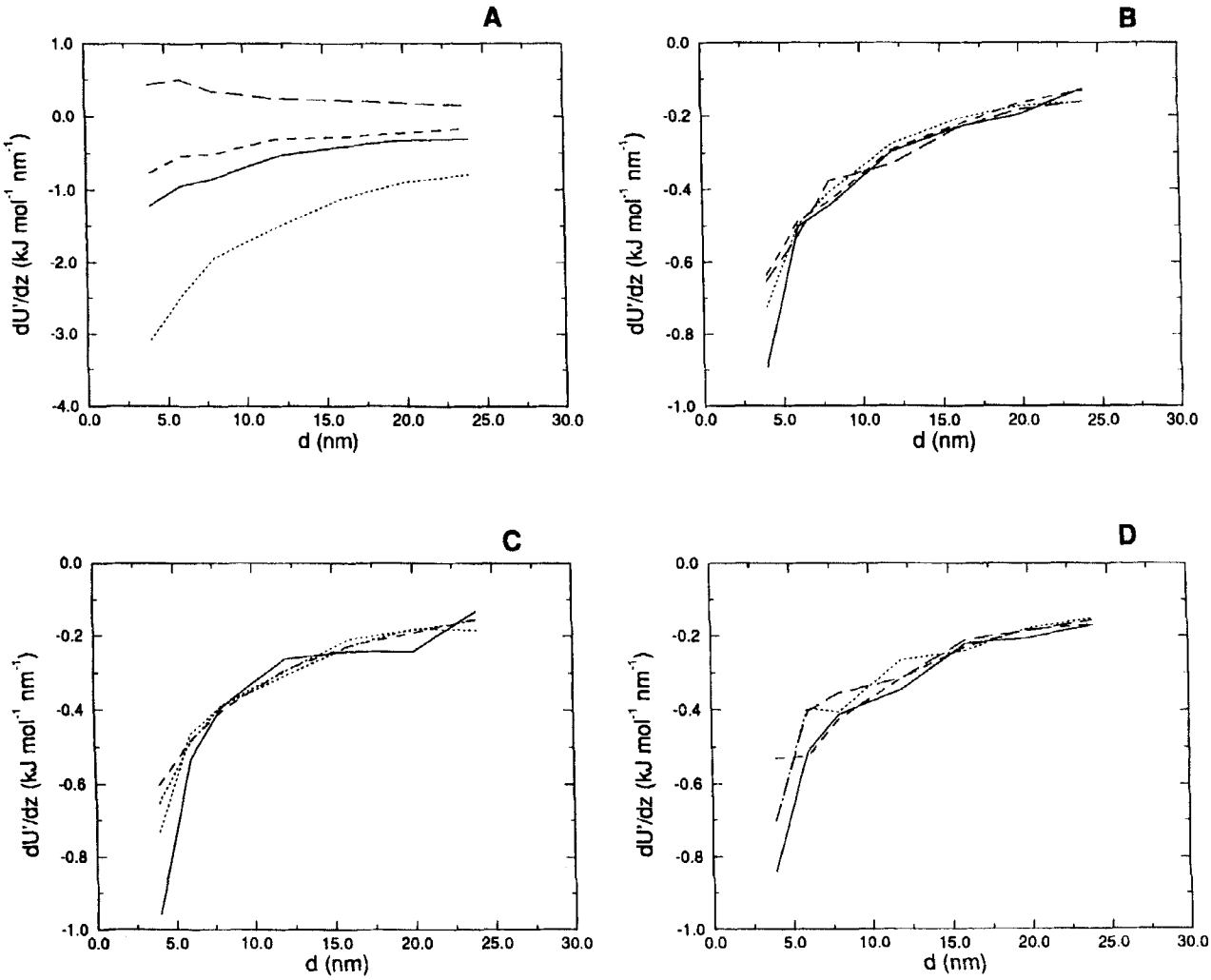

FIGURE 4. $\langle \partial U'/\partial z \rangle$ (in the plots the symbol $\langle dU'/dz \rangle$ is used) (in kJ mol^{-1} nm^{-1}) versus the distance d (nm) of the macromolecular center of mass from the charged surface. (A) n172k (solid line), t45k (dotted), wild-type (dashed), r17e (long-dashed). (B) a1 (solid line), a2 (dotted), a3 (dashed), a5 (long-dashed). (C) b1 (solid line), b2 (dotted), b3 (dashed), b5 (long-dashed). (D) a1 (solid line), c2 (dotted), c3 (dashed), c5 (long-dashed).

culate points between the measured $\langle \partial U'/\partial z \rangle$ was employed to determine the integral in eq. (8) from $d = 4$ to $d = 24$ nm. For the set of proteins with a total charge of $+1$ (see Table III), a smoothing operation between $d = 12$ to 24 nm was employed to ensure that possible errors in the approximation or uncertainties in the measured $\langle \partial U'/\partial z \rangle$ did not significantly affect the value of the adsorption free energy.

Table IV lists the final results for all proteins. The free energy of adsorption is in the range $+10$ (4 kT) and -50 (-20 kT) kJ mol^{-1} for the cases studied in which the wild-type had a value of -8.86 (-3.6 kT) kJ mol^{-1}. Variant t45k displayed the highest affinity of all variants for the charged surface.

Figure 5 shows the correlation between strength of the electric moment and $\langle \partial U'/\partial z \rangle$ for $d = 4$ and 6 over the variants with a total charge of $+1$ including the wild-type. The electric moment was defined[41] as $\mu = \Sigma q_i(\mathbf{r}_i - \mathbf{R})$ where $\mathbf{r}_i - \mathbf{R}$ is the position vector of charge i with respect to the center of mass $\mathbf{R}$ of the protein. The strength μ of the electric moment is simply the length of vector μ. The term "electric moment" is preferred over "dipole moment" because the latter is generally applied to a molecule with zero total charge (e.g., water).[41] Clearly, close to the surface ($d = 4$ nm, Fig. 5A), the value of $\langle \partial U'/\partial z \rangle$ varies linearly with the strength of the electric moment. For this distance, there were no large differences between the averages over the first $2.5 \times 10^{+6}$ and the

TABLE IV.
Numerical Values for Helmholtz Free Energy of Adsorption (ΔA) in kJ mol^{-1}.

Protein	ΔA (kJ mol^{-1})
wt	−8.86
n172k	−14.44
t45k	−41.21
r17e	7.52
a1	−8.73
a2	−8.54
a3	−8.51
a5	−8.49
b1	−8.73
b2	−8.45
b3	−8.43
b5	−8.45
c1	−8.66
c2	−8.32
c3	−8.51
c5	−8.24

[a] The first column refers to the protein code, the second lists free energies.

second $2.5 \times 10^{+6}$ Monte Carlo steps for all +1 variants except for wild-type cutinase and variant a3 for which the simulations were prolonged to $7.5 \times 10^{+6}$ steps, resulting in stable averages. Prolongation of the simulations for the variants a2, a5, b2, and c2 with similar electric moments as the wild-type up to $7.5 \times 10^{+6}$ steps yielded differences between $\langle \partial U'/\partial z \rangle$ over the first $5 \times 10^{+6}$ Monte Carlo steps and over the full run within the error of $\langle \partial U'/\partial z \rangle$ (see *Error Analysis* subsection).

At a distance of 6 nm, the value of $\langle \partial U'/\partial z \rangle$ differs more among the variants and correlation with the electric moment is reduced (Fig. 5B). At a distance of 8 nm (results not shown), all variants display similar values for $\langle \partial U'/\partial z \rangle$, except for the wild-type, which experiences a stronger attractive force. The simulations for all b-variants, c3, and the wild-type were prolonged to $7.5 \times 10^{+6}$ steps. There were almost no differences between the averages for $\langle \partial U'/\partial z \rangle$ over the first $5 \times 10^{+6}$ steps and the full run.

Figure 6A shows the correlation between the total charge q of all variants and the adsorption free energy for which a clear linear correlation is established. The adsorption free energy varies to a much less degree with the strength of the electric moment for those variants which have a total charge of +1 (Fig. 6B). Figure 6C shows the correlation between the strength Q of the quadrupole moment of those variants with the same total

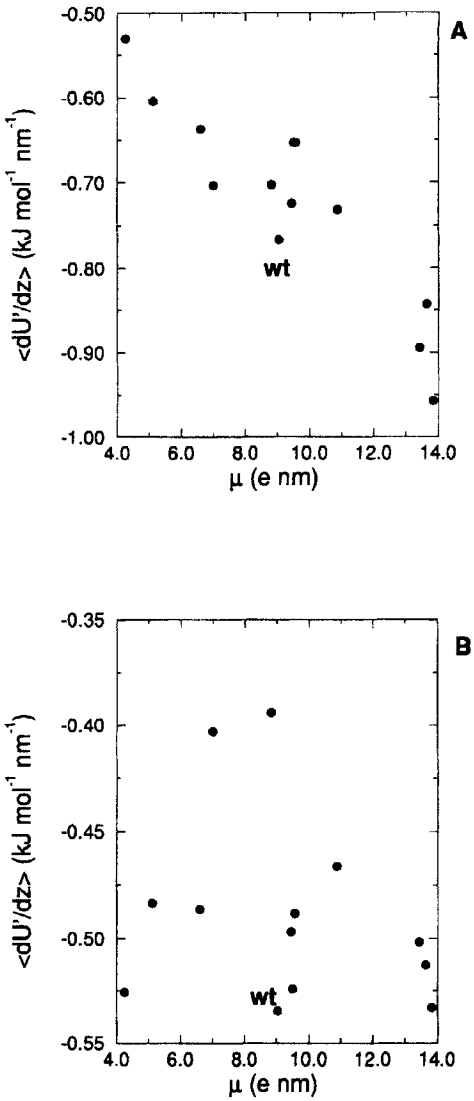

FIGURE 5. $\langle \partial U'/\partial z \rangle$ (in the plots the symbol $\langle dU'/dz \rangle$ is used) (in kJ mole^{-1} nm^{-1}) versus the strength of the electric moment μ (e nm). Results are shown for the variants with a total charge of +1. (A) At a distance of 4 nm between the protein center of mass and the charged surface. (B) At a distance of 6 nm between the protein center of mass and the charged surface.

charge and a similar electric moment as the wild-type (wild-type variants a2, a5, b2, and c2; see Table III) from which it is seen that the wild-type is at a minimum of free energy. The quadrupole tensor $\mathbf{Q}$ is defined[41] as $\frac{1}{2}\Sigma_i q_i(3\mathbf{r}_j\mathbf{r}_i - r_i^2\mathbf{I})$, where $\mathbf{r}_i$ is the position vector of charge q_i relative to the protein's center of mass and $\mathbf{I}$ is the unit tensor. The strength of the quadrupole moment is determined by $Q^2 = \Sigma_{i=1}^{3}\Sigma_{j=1}^{3}(2Q_{ij}^2 - Q_{ii}Q_{jj})$, where Q_{ij} refers to a tensor element. Figure 6D shows a

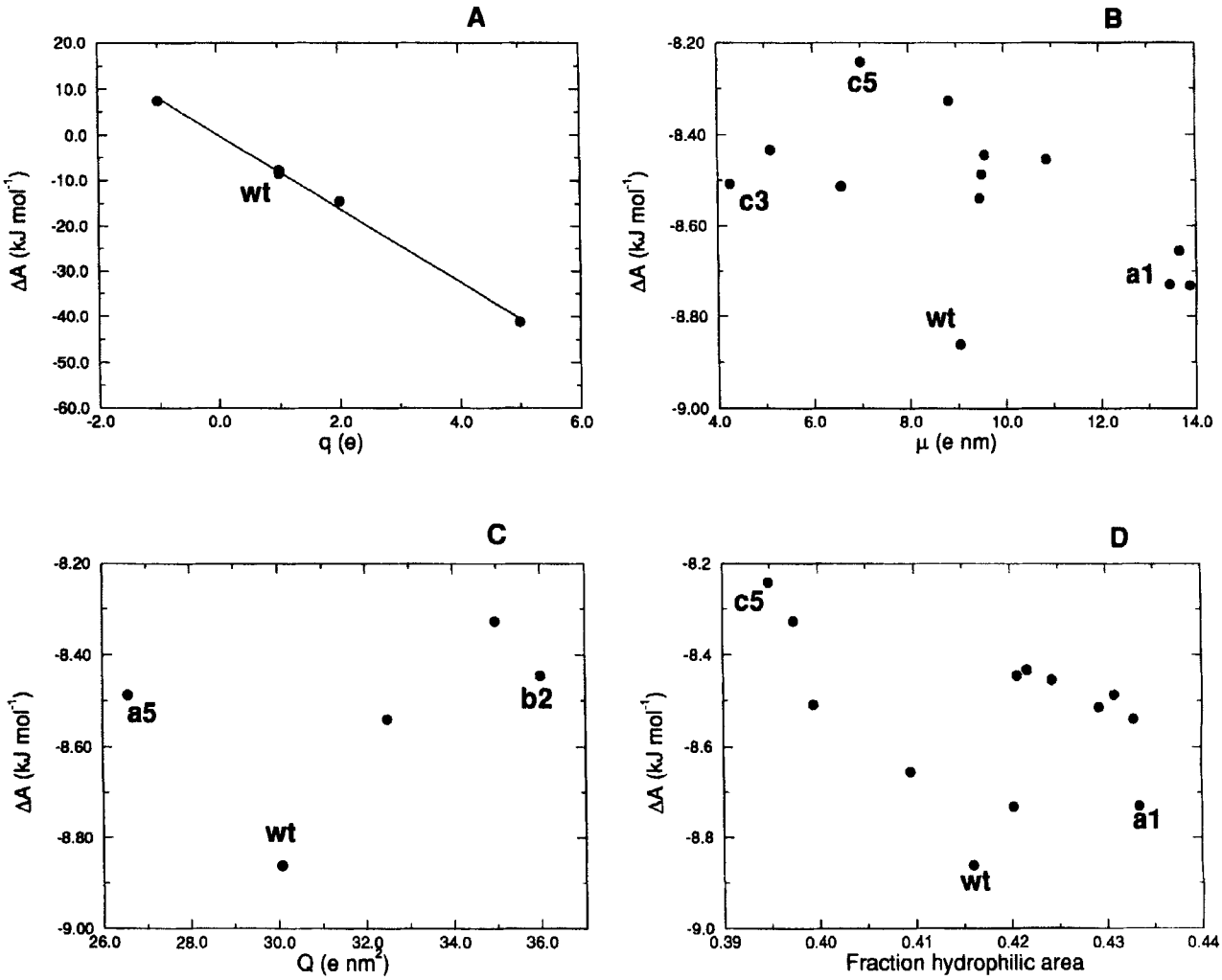

FIGURE 6. Free energy of adsorption, ΔA (in kJ mol^{-1}). (A) Versus total charge q (in e) for all variants. (B) Versus the strength of the electric moment μ (e nm) (defined in the text) for all variants with a total charge of +1 (wild-type, a1, a2, a3, a5, b1, b2, b3, b5, c2, c2, c3, and c5; see Table III). (C) Versus the strength of the quadrupole moment Q (e nm^2) (defined in the text) for those variants with the same total charge and a similar electric moment as wild-type cutinase (a2, a5, b2, and c2 and the wild-type; see Table III). (D) Versus the fraction of hydrophilic macromolecular surface area for all variants with a total charge of +1 (wild-type, a1, a2, a3, a5, b1, b2, b3, b5, c1, c2, c3, and c5; see Table III).

plot of the fraction of hydrophilic macromolecular surface area versus the free energy of adsorption for all +1 variants from which it is seen that the wild-type is at a minimum of free energy. The fraction of hydrophilic surface area was determined from the solvent accessible surface[42] and atomic solvation parameters (ASP).[43] If the ASP of a particular atom was negative, it was assigned to be a hydrophilic atom. For the variants, it varied between 39.5% and 43.3% (Table III). The most striking result from Figure 6 and Table IV is that the wild-type has the largest free energy of adsorption of all +1 variants although its electric moment is not the largest. The use of interpolation schemes other than that employed here to determine the free energy of adsorption yielded similar results as presented in Figure 6.

Orientation Distributions

Although it is possible to measure experimentally the orientation of a protein with respect to the surface,[44] no systematic study known by the authors has been performed on the relationship between electric moment and the orientation of the protein. In this work, the distribution $f(\cos\phi)$ of the cosine of the angle ϕ between the electric moment vector and the vector (0, 0, 1) (vector par-

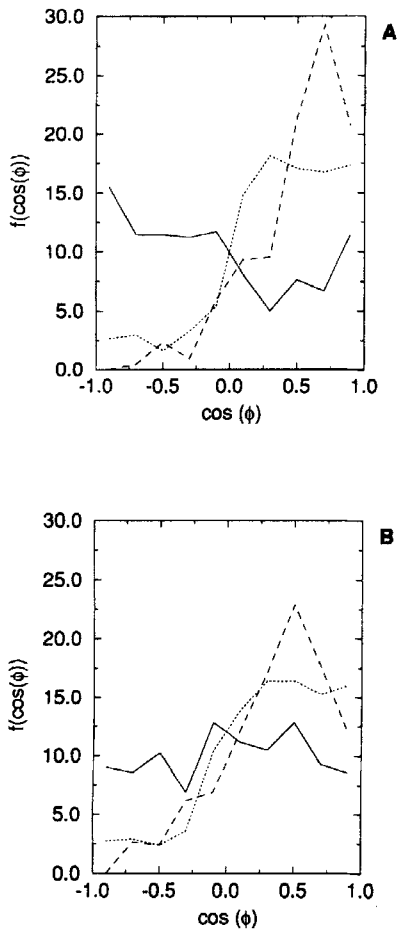

FIGURE 7. Orientation distributions $f(\cos(\phi))$ of the cosine of the angle ϕ normalized between 0 and 100. Results are shown for the variants c3 (solid lines), wild-type (dotted lines), and a1 (dashed lines). Note that the charged boundary is perpendicular to the positive z-axis and parallel to the xy-plane. (A) Distribution of the electric moment where ϕ refers to the angle between the electric moment and the positive z-axis. (B) Distribution of the vector from the C_α of Ile-115 to the C_α of Ser-120 where ϕ refers to the angle between this vector and the positive z-axis.

allel to the positive z-axis) was examined. A vanishing cosine indicates a nearly parallel orientation of the electric moment with respect to the surface. Because cutinase has an active site consisting of the residue triad Ser-120, His-188, and Asp-175, it is practical to consider the relative orientation of the active site with respect to the charged surface. Active site positioning was characterized by a vector from the C_α of Ile-115 to the C_α of Ser-120 such that it is parallel to the central β-sheet of the protein and points toward the crevice in which Ser-120 is located (see Fig. 1). If $\cos\phi$ is equal to 1, then the active site faces the surface. Figure 7 shows exemplary results on the orientation distributions for variant c3, wild-type cutinase, and variant a1, respectively, for a distance of 4 nm of their center of mass from the charged surface.

Close to the surface (d = 4 nm), the protein orientation of a protein just follows the strength of the electric moment. Therefore, variant a1 shows a particulary strong orientational preference toward the surface and c3 does not. At this distance the screening by ions is not sufficient to reduce the orientation effect of the electric field from the surface although there remains 1.0–1.5 nm between the protein surface and the charged surface. On the other hand, the field is apparently not strong enough to orient variant c3, which has the smallest electric moment of all variants. The orientation distribution of the wild-type is broader in comparison with variant a1. These results may be compared with Figure 5A. At a distance of 6 nm (results not shown), the ion screening is such that any orientational preference is small or completely lost. However, the wild-type still displays an orientational preference toward the surface at 6 nm and also at 8 nm (data not shown). This may be related to the fact that the wild-type experiences a relatively strong attractive mean force from the surface at both distances a shown earlier. No orientational preferences of importance were observed for larger distances.

Error Analysis

An example is given for wild-type cutinase at a center of mass distance of 6 nm from the surface in Figure 8. The value of the statistical inefficiency, s, is about 60 where, for its determination, a straight line was fitted through all points with $(1/l_b)^{-1} <$ 0.01 (Fig. 8A). The error in the mean $\langle \partial U'/\partial z \rangle$ (its value was -0.53 at 6 nm) is about 0.015 ($\sim$ 3%) which is quite typical of the accuracy obtainable for computer simulations.[35] Figure 8C shows the error as a function of the block length. Figure 8B shows the autocorrelation function $c(\tau)$ (τ is "time," but here it obviously refers to "number of states," a state is characterized by a particular set of {r}, φ, θ, and ψ [see eq. (4)] of the system) of the $\partial U'/\partial z$ data series, which indicates the existence of a correlation up to 140 states. Using s = 140 results in the same level of error (Fig. 8C). It is concluded that the accuracy of the simulations is good.

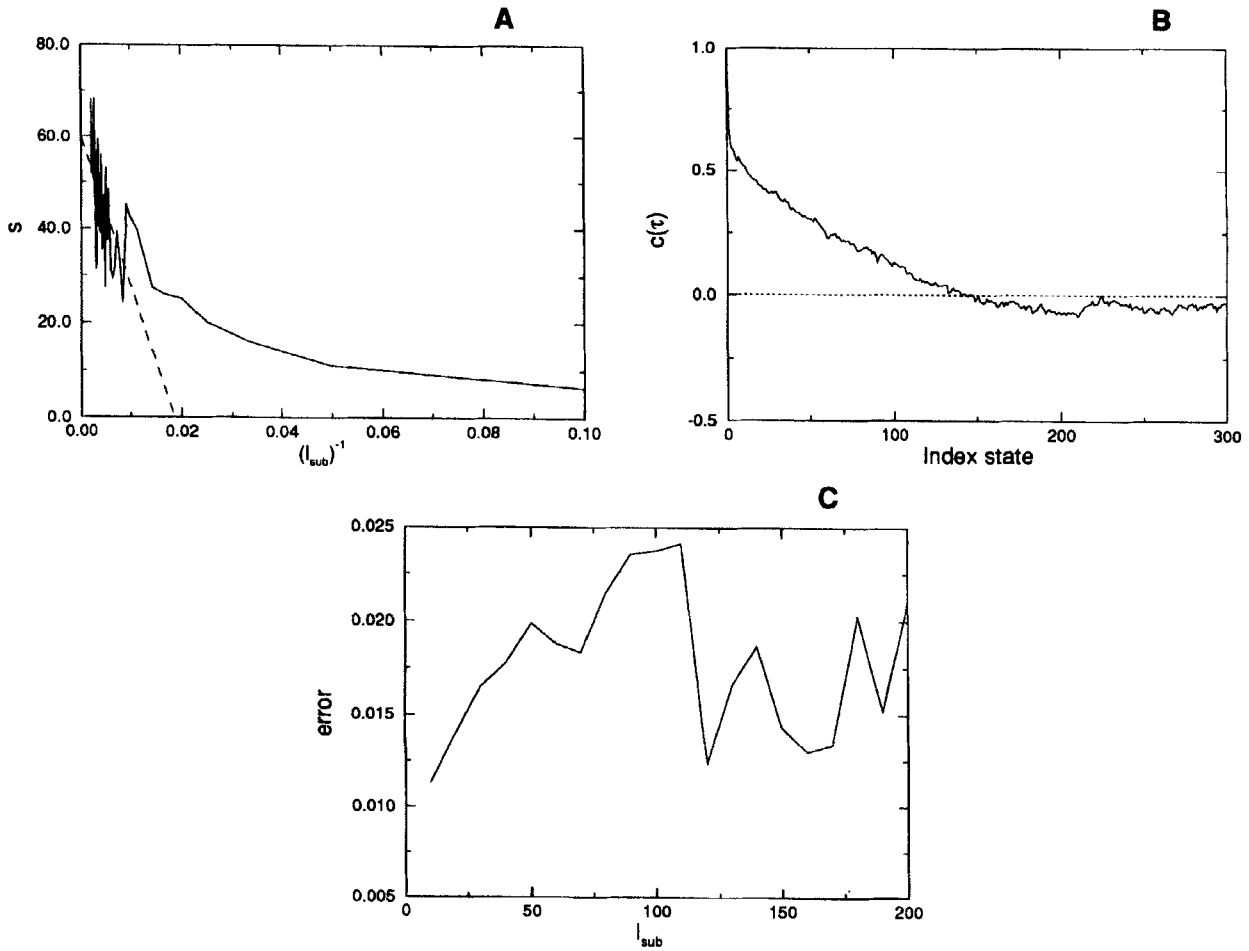

FIGURE 8. Error analysis of $\langle \partial U'/\partial z \rangle$ (in the plots the symbol $\langle dU'/dz \rangle$ is used) for the wild-type cutinase at a distance of 6 nm between the center of mass and the charged surface. (A) Statistical inefficiency s versus the inverse of the block length l_{sub}. (B) Autocorrelation function $c(\tau)$ of data series $\partial U'/\partial z$. The symbol τ refers to "time," but obviously here it refers to "numbers of states." A state is characterized by a particular set of $\{\mathbf{r}\}$, φ, θ, and ψ [see eq. (4)] of the system. The x-axis label, "Index state," refers to the states. (C) Error as a function of block length l_{sub}.

Discussion

This work has focused on a simulation model for the adsorption of proteins onto charged surfaces. The current model principally considered electrostatic aspects, but inclusion of many other types of interactions are clearly possible. The observed linear correlation [eq. (16)] between the total charge of a protein and its adsorption free energy (Fig. 6A) confirms the consistency of the simulations and is in agreement with experimental data.[4-6] The degree of variation in the observed values for the mean force ($\langle \partial U'/\partial z \rangle$), experienced by the variants at several positions with respect to the surface, indicates that the simulation model is sensitive to changes in the charge distribution of the proteins. The small errors in the values of $\langle \partial U'/\partial z \rangle$ point to the significance of the differences between the mean forces experienced by the variants. The accuracy is further confirmed by the persistence of the values of $\langle \partial U'/\partial z \rangle$ after prolongating the simulations. Also, since the total charge of the protein is the determining factor for the adsorption free energy, small differences should be expected between several variants with the same total charge.

From the results for the variants with total charge of $+1$, it can be concluded that the net contribution of the electric moment to the free energy of adsorption is a second-order one for proteins; a twofold increase of the strength of the

electric moment does not lead to a twofold better adsorption. The total charge is a major determinant. A similar conclusion was drawn by Roth and Lenhoff[14] based on a continuum approach, although, as noted by Roush et al.,[15] a patch of charges can also be the dominant factor in determining the free energy of adsorption. At short distances (d = 4 nm), the electric moment almost completely determines the *differences* between the *mean forces* acting on the protein and therefore its orientation. The electric moment is important for a proper orientation with respect to the surface of the protein such that its active site always faces the surface. It is striking that the wild-type is properly oriented already at distances of 6 and 8 nm, whereas all other variants do not display any preferential orientation at these distances. For enzyme activity, this may be of importance. If the total charge of the protein has the same sign as the charged surface (variant r17e; Fig. 4A), the interaction with the surface is mainly repulsive, but there seems to be a maximum in the $\langle \partial U'/\partial z \rangle$ curve which may be due to a favorable interaction of the electric moment (comparable with the wild-type) of this variant with the surface.

Coadsorption of ions at the charged surface was included by using an explicit ion representation. Protein–ion correlations may also have influence on the adsorption behavior of a protein. Whenever two oppositely charged surfaces approach each other, the release of ions from both surfaces could be considered a driving force for binding.[45] Monte Carlo simulations of several cutinase variants in 0.1 M NaCl solution (results not given) showed that ions can enter the outer regions of the protein up to within 0.5 nm of the protein surface, for comparison, the radius of wild-type cutinase is about 2.5 nm. In fact, the release of these aspecific bound ions is an entropic effect which may be compared with the hydrophobic effect in the protein folding process.[45] Such effects were considered in this work by plotting the fraction of hydrophilic protein surface area versus the free energy of adsorption (Fig. 6D). Although much less convincing in comparison to the correlation between electric moment and free energy (Figs. 5A and 6B), it is possible that an increasing fraction of hydrophilic area (or equivalently a decreasing fraction of hydrophobic area) results in an increasing affinity of the protein for the surface. Nevertheless, this effect remains second order, because the total charge of the protein is a major determinant.

The fact that wild-type cutinase displays the highest affinity for the charged surface of all variants with same total charge as the wild-type (+1) suggests that the charge distribution of wild-type cutinase is optimal for adsorption at charged surfaces (at least within the limits of the present model). All newly designed variants (with the same total charge) destroy this optimality. The orientation distribution of the wild-type (Fig. 7) is also quite effective: the protein always faces the surfaces, even at larger distances. In Figure 6D, the correlation between the strength of the quadrupole moment and adsorption free energy is shown for those variants with the same total charge of +1 and a similar electric moment as the wild-type. Although the number of points in the graph is rather limited, the data suggest an optimum for the wild-type. The quadrupose strength of wild-type cutinase is one of the smallest. Only the variants a5, c3, and c5 have smaller quadrupole moments.

Clearly, the adsorption model is not complete. Inclusion of polarization effects could effect the results. It is important to include the hydrophobic effect[5] through explicit water molecules, albeit requiring excessively long simulations which can be mitigated by use of an effective potential. Such a potential can be derived from a computer simulation of a protein (or several proteins) in water (without ions) next to an (un)charged surface and corresponds to a potential of mean force.[25] Some work in this direction was done by Alosi et al.[46,47] who performed Monte Carlo simulations of water with and without nonpolar monomeric solute molecules next to a (un)charged wall. An additional aspect requiring consideration in the current model is the pH of the solution.[6,20] The charge distribution of any protein is a fluctuating quantity and so is its total charge which is a major determinat for adsorption. The protonation state of titrating sites of surface residues of proteins strongly depends on local electrostatic effects such as interactions between titrating sites but also the interaction with the charged surface (of which the charge distribution generally depends on the pH also). The latter may lead to additional shifts in the pK_a (K_a is the acid dissociation constant) of titrating sites present at the interface between protein and (charged) surface and, therefore, leads to changes in the overall charge distribution on the protein and the surface. In addition, the pH next to a charged surface may be different from that of the bulk solution.[48] The effect of the pH can be in-

cluded as an additional degree of freedom in the system such that protons are being distributed over titrating sites.[49] In general, the binding of ions (including protons) at sites of the protein and the surface will have impact on the adsorption behavior of a protein. The current model only includes aspecific binding of ions; namely, in the double layer next to the charged surface and around the protein.

This work demonstrates that it is feasible to perform adsorption simulations with atomic detail for the protein and the solvent. The amount of CPU time for 250,000 steps on a Silocon Graphics workstation with an R8000 processor running at 75 MHz took about 8 hours. To determine a full free energy profile (such as in Fig. 4), about 3 weeks of CPU time were required per variant. The computer program was written in C++ computer language.[50]

Conclusions

A consistent simulation model for the study of adsorption of proteins onto charged surfaces has been developed. The model should prove useful in the design and engineering of proteins, especially through residue substitutions, whose effect can be tested by computer simulations. The method is sensitive to changes in the charge distribution of proteins, and the accuracy has been shown to be good. Also, basic principles of adsorption can be studied.

The basic model can be extended to include asymmetric ions, atomic detail at the surface, small organic molecules, surfactants, effect of pH, and so forth. Currently, tests are underway to include pH and hydrophobic effects. Both will address the so-called charge regulation effect upon binding of a protein to a surface, which also involves the coadsorption of ions at the charged surface and the adsorbing protein.

The present model applied to cutinase and its variants showed that the total charge is the major determinat for adsorption, whereas the electric moment (and possibly the hydrophilic fraction of protein surface area) has a second-order effect. The wild-type displayed, among all variants with the same total charge, the highest affinity for the charged surface, suggesting that the charge distribution of cutinase is optimal for binding at lipid surfaces. The combination of efficient adsorption and proper orientation also suggests that the native enzyme will show the highest catalytic activity of all the +1 variants. Though increasing the total charge of the protein clearly leads to a better adsorption, it may well be more difficult to remove the protein from the surface or adsorption site, and the higher charge on the protein may lead to destabilization of the protein's folded state with resultant negative effects on the catalytic efficiency.

Acknowledgments

A. H. J. thanks the Unilever Corporation (The Netherlands) for providing a long-term postdoctoral fellowship. The authors are grateful to Hans Peters and Marco Hoppenbrouwers of Unilever for preparing the structures of cutinase variants by computer modeling. The authors also thank Michael Nilges and Philip Lijnzaad of EMBL Heidelberg for helpful discussions, and Jaap Heringa and Frank Eisenhaber, also of EMBL, for a careful reading of the manuscript and stimulating discussions.

Appendix A

The purpose of this Appendix is to derive eq. (13). The electric potential, $\phi(\mathbf{r}_0)$, at $\mathbf{r}_0$ of a square Σ (area L^2) carrying a uniform surface charge density σ_P is[28]:

$$\phi(\mathbf{r}_0) = \int_\Sigma \frac{\sigma_P}{4\pi\varepsilon\varepsilon_0|\mathbf{r} - \mathbf{r}_0|}\, d\sigma \qquad (21)$$

Here, $\mathbf{r}$ is a point on Σ, $\int_\Sigma$ is a surface integral and the reference point is taken at infinity.

If the point $\mathbf{r}_0$ is located at a distance r_0 just below the center of the square, then eq. (21) can be rewritten as, using polar coordinates r and φ,

$$\phi(r_0) = \frac{\sigma_P}{4\pi\varepsilon\varepsilon_0} 8 \int_0^{1/4\pi} \int_0^{p} \frac{1}{(r^2 + r_0^2)^{1/2}} r\, dr\, d\varphi \qquad (22)$$

where $p = L/(2\cos\varphi)$. The factor 8 arises from symmetry, since the square can be divided into eight equivalent parts. In eq. (22), the integration over r is straightforward yielding:

$$\phi(r_0) = \frac{2\sigma_P}{\pi\varepsilon\varepsilon_0} \int_0^{1/4\pi} \left(\frac{\frac{1}{4}L^2 + r_0^2\cos^2\varphi}{\cos^2\varphi} \right)^{1/2} d\varphi - \frac{\sigma_P}{2\varepsilon\varepsilon_0} r_0 \qquad (23)$$

The electric potential, $\psi(r_0)$, of an infinite plane at a distance r_0, carrying a uniform surface charge density σ_P is[28]:

$$\psi(r_0) = -\frac{\sigma_P}{2\varepsilon\varepsilon_0} r_0 \tag{24}$$

The reference point for this equation is taken at $r_0 = 0$ (i.e., on the surface).

After choosing a common reference point (e.g., at $r_0 = 0$), substracting eq. (23) from eq. (24) leads to the desired result for the interaction $u_i^P(d_i)$ of a single charge q_i at distance d_i (which replaces r_0 in the previous equations) to the virtual plane P; namely:

$$u_i^P(d_i) = -q_i \frac{2\sigma_P}{\pi\varepsilon\varepsilon_0} \times \int_0^{1/4\pi} \left(\frac{\frac{1}{4}L^2 + d_i^2\cos^2\varphi}{\cos^2\varphi} \right)^{1/2} d\varphi + C \tag{25}$$

The constant C depends on the choice of the reference point, although the actual choice is of no relevance, since only energy differences are considered in the Monte Carlo method. For the sake of completeness, $C = \sigma_P/\pi\varepsilon\varepsilon_0\, L(\ln\tan\frac{3}{8}\pi)$, if the reference point is at the surface. The integral can be solved using, for example, Simpson's rule of integration.[39,40] Eq. (25) will result in a damping effect on the electric field of a virtual plane when d_i approaches zero. For large values of d_i, $\frac{1}{4}L^2 + d_i^2\cos^2\varphi$ approaches $d_i^2\cos^2\varphi$ and $u_i^P(d_i)/q_i$ becomes similar to eq. (24). The same is true in the limit of zero L.

Appendix B

The purpose of this Appendix is derive an expression for $\partial U'/\partial z$ which, in the first instance, is given as:

$$\frac{\partial U'}{\partial z} = \sum_{l=1}^{n_{cg}} \frac{\partial u_l^B(d_l^B)}{\partial d_l^B} \frac{\partial d_l^B}{\partial z} + \sum_{j=1}^{M} \sum_{l=1}^{n_{cg}} \frac{\partial u_l^{P_j}(d_l^{P_j})}{\partial d_l^{P_j}} \frac{\partial d_l^{P_j}}{\partial z} + \sum_{k=1}^{n_{ions}} \sum_{l=1}^{n_{cg}} \frac{\partial u_{kl}^{PM}(r_{kl})}{\partial r_{kl}} \frac{\partial r_{kl}}{\partial z}. \tag{26}$$

An expression for $\partial d/\partial z$ is first derived (d is the distance to the surface or to a virtual plane). Here, z is the z-component of $\mathbf{R}$, the position vector of the center of mass of the macromolecule located along the z-axis. The position vector, $\mathbf{r}_l$, of atom l in the macromolecule can be expressed as:

$$\mathbf{r}_l(z, \theta, \psi) = \mathbf{R}(z) + \mathbf{r}_l^{CM}(\theta, \psi) \tag{27}$$

where the vector $\mathbf{r}_l^{CM}$ corresponds to the position of atom l with respect to the macromolecular center of mass (CM) of the macromolecule. The components of this vector depend only on the orientation of the macromolecule and not on the position of its center of mass.

Now, for d_l, one can write:

$$d_l = a - z_l = a - z - z_l^{CM} \tag{28}$$

The constant a corresponds to the location of the charged boundary or a virtual plane.

From eqs. (27) and (28), it can be shown that:

$$\frac{\partial d_l}{\partial z} = -1 \tag{29}$$

Similarly, an expression for $\partial r_{kl}/\partial z$ can be derived. Here:

$$r_{kl} = \left[(x_l - x_k)^2 + (y_l - y_k)^2 + (z + z_l^{CM} - z_k)^2 \right]^{1/2} \tag{30}$$

and:

$$\frac{\partial r_{kl}}{\partial z} = \frac{z_l - z_k}{r_{kl}} \tag{31}$$

From eq. (24), it can be shown that:

$$\frac{\partial u_l^B}{\partial d_l^B} = -q_l \frac{\sigma_B}{2\varepsilon\varepsilon_0} \tag{32}$$

and

$$\frac{\partial u_{kl}^{PM}}{\partial r_{kl}} = -u_{kl}^{PM} \frac{1}{r_{kl}} \tag{33}$$

The second term in eq. (26) is more complex, since it involves a differentiation below the integral sign [eq. (25)]. Since the functions involved

are all continuous, the differentiation can be performed[51] such that:

$$\frac{\partial u_l^{P_j}}{\partial d_l^{P_j}} = -q_l \frac{2\sigma_{P_j}}{\pi\varepsilon\varepsilon_0} d_l^{P_j} \times \int_0^{\frac{1}{4}\pi} \left\{ \frac{\frac{1}{4}l^2 + (d_l^{P_k})^2 \cos^2\varphi}{\cos^2\varphi} \right\}^{-1/2} d\varphi \quad (34)$$

The integral is standard and can be solved analytically.[52] Combining eqs. (26) to (34), the desired result becomes:

$$\frac{\partial U'}{\partial z} = \frac{\sigma_B}{2\varepsilon\varepsilon_0} Q + \sum_{j=1}^{M} \sum_{l=1}^{n_{cg}} q_l \frac{2\sigma_{P_j}}{\pi\varepsilon\varepsilon_0} \times \sin^{-1}\left(\frac{(d_l^{P_j})^2}{\frac{1}{4}l^2 + (d_l^{P_j})^2} \frac{1}{2}\sqrt{2} \right) - \sum_{k=1}^{n_{ions}} \sum_{l=1}^{n_{cg}} u_{kl}^{PM}(r_{kl}) \frac{z_k - z_l}{r_{kl}^2}. \quad (35)$$

References

1. Z. S. Derewenda, *Adv. Prot. Chem.*, **45**, 1 (1994).
2. L. Stryer, *Biochemistry*, 2nd ed., W. H. Freeman & Co., New York, 1981.
3. J. Sjollema, Ph.D. thesis, University of Groningen, The Netherlands, 1990.
4. (a) T. M. Lohman, P. L. DeHaseth, and M. T. Record, *Biophys. Chem.*, **8** 281 (1978); (b) W. Norde, *Adv. Coll. Interf. Sci.*, **25**, 267 (1986).
5. (a) F. MacRitchie, *Adv. Prot. Chem.*, **32**, 283 (1978); (b) W. Norde and J. Lyklema, *J. Biomater. Sci. Polym. Ed.*, **2**, 183 (1991).
6. M. B. Sankaram and D. Marsh, In *Protein–Lipid Interactions*, A. Watts, Ed., Elsevier, Amsterdam, 1993, p. 127.
7. T. Heinburg and D. Marsh, *Biophys. J.*, **68**, 536 (1995).
8. T. E. Creighton, *Proteins: Structures and Molecular Properties*, W. H. Freeman & Co., New York, 1993.
9. S. Duinhoven, Ph.D. thesis, Agricultural University Wageningen, The Netherlands, 1992.
10. R. Aveyard and D. A. Haydon, *An Introduction to the Principles of Surface Chemistry*, Cambridge University Press, Cambridge, 1973.
11. J. N. Israelachvili, *Intermolecular and Surface forces. With applications to Colloidal and Biological Systems*, Academic Press, New York, 1985.
12. K. E. Forsten, R. E. Kozack, D. A. Lauffenburger, and S. Subramaniam, *J. Phys. Chem.*, **98**, 5580 (1994).
13. B. J. Yoon and A. M. Lenhoff, *J. Phys. Chem.*, **96**, 3130 (1992).
14. C. M. Roth and A. M. Lenhoff, *Langmuir*, **9**, 962 (1993).
15. (a) D. J. Roush, D. S. Gill, and R. C. Willson, *Biophys. J.*, **66**, 1290 (1994); (b) D. S. Gill, D. J. Roush, and R. C. Wilson, *J. Chromatogr. A*, **704**, 339 (1995).
16. T. L. Hill, *An Introduction to Statistical Thermodynamics*, Dover, New York, 1986.
17. D. A. McQuarrie, *Statistical Mechanics*, Harper and Row, New York, 1976.
18. E. D. Getzoff, D. E. Cabelli, C. L. Fisher, H. E. Parge, M. S. Viezzoli, L. Banci, and R. A. Hallewell, *Nature*, **358**, 347 (1992).
19. R. J. Bacquet, J. A. McCammon, and S. A. Allison, *J. Phys. Chem.*, **92**, 7134 (1988).
20. J. G. E. M. Fraaije, W. Norde, and J. Lyklema, *Biophys. Chem.*, **41**, 263 (1991).
21. (a) B. Jayaram, K. A. Sharp, and B. Honig, *Biopolymers*, **28**, 975 (1989); (b) B. Jayaram, S. Swaminathan, D. L. Beveridge, K. Sharp, and B. Honig, *J. Phys. Chem.*, **23**, 3156 (1990); (c) B. Jayaram, S. Swaminathan, D. L. Beveridge, K. Sharp, and B. Honig, *Macromolecules*, **23**, 3156 (1990); (d) B. Jayaram, F. M. DiCapus, and D. L. Beveridge, *J. Am. Chem. Soc.*, **113**, 5211 (1991).
22. C. Martinez, P. De Geus, M. Lauwereys, G. Matthyssens, and C. Cambilleau, *Nature*, **356**, 615 (1992).
23. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).
24. G. M. Torrie and J. P. Valleau, *J. Chem. Phys.*, **73** 5807 (1980).
25. T. L. Hill, *Statistical Mechanics. Principles and Selected Applications*, Dover, New York, 1987.
26. H. Goldstein, *Classical Mechanics*, 2nd ed., Addison-Wesley, Amsterdam, 1980.
27. P. King, In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, Eds., Escom. Leiden, 1993, p. 267.
28. D. J. Griffiths, *Introduction to Electroydynamics*, Prentice-Hall, Englewood Cliffs, NJ, 1981.
29. (a) M. K. Gilson and B. H. Honig, *Biopolymers*, **25**, 2097 (1986); (b) M. K. Gilson and B. H. Honig, *Nature*, **330**, 84 (1987); (c) M. K. Gilson and B. H. Honig, *Prot. Struct. Function Genet.*, **3**, 32 (1988).
30. A. H. Juffer, E. F. F. Botta, B. A. M. van Keulen, A. van der Ploeg, and H. J. C. Berendsen, *J. Comput. Phys.*, **97**, 141 (1991).
31. J. Antosiewics, J. A. McCammon, and M. K. Gilson, *J. Mol. Biol.*, **238**, 415 (1994).
32. D-M. Soumpasis, *Proc. Natl. Acad. Sci. USA*, **81**, 5116 (1984).
33. B. Jönsson and B. Svensson, In *Computer Simulation of Biomolecular Systems*, Vol. 2, W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, Eds., Escom, Leiden, 1993.
34. P. E. Smith, R. M. Brunne, A. E. Mark, and W. F. van Gunsteren, *J. Phys. Chem.*, **97**, 2009 (1993).
35. M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids*, Clarendon Press, Oxford, 1987.
36. D. N. Card and J. P. Valeau, *J. Chem. Phys.*, **52**, 6232 (1970).
37. W. F. van Gusteren and H. J. C. Berendsen, *Angew. Chem. Ed. Engl.*, **29**, 992 (1990).
38. InsightII, Biosym Technologies, Inc., Sand Diego, CA, USA, Version 2.3.5, 1994.
39. F. B. Hildebrand, *Introduction to Numerical Analysis*, 2nd ed., Dover, New York, 1987.

40. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes. The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1986.
41. C. J. F. Böttcher, *Theory of Electric Polarization*, Vol. 1, 2nd ed., Eslevier, Amsterdam, 1973.
42. F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf, *J. Comput. Chem.*, **16**, 273 (1995).
43. D. Eisenberg and A. D. McLachlan, *Nature*, **319**, 199 (1986).
44. J. G. E. M. Fraaije, J. M. Kleijn, M. van der Graaf, and J. C. Dijt, *Biophys. J.*, **57**, 965 (1990).
45. M. T. Record Jr., C. F. Anderson, and T. M. Lohman, *Q. Rev. Biophys.*, **11**, 103 (1978).
46. G. Aloisi, M. L. Foresti, R. Guidelli, and P. Barnes, *J. Chem. Phys.*, **91**, 5592 (1989).
47. G. Aloisi and P. Guidelli, *J. Chem. Phys.*, **95**, 3679 (1991).
48. F. G. van der Goot, J. M. Gonzáles-Mañas, J. H. Lakey, and F. Pattus, *Nature*, **354**, 408 (1991).
49. P. Beroza, D. R. Fredkin, M. Y. Okamura, and G. Feher, *Proc. Natl. Acad. Sci. USA*, **88**, 5804 (1991).
50. S. B. Lippman, *C++ Primer*, 2nd ed., Addison-Wesley, New York, 1991.
51. T. M. Apostol, *Calculus*, Vol. II, 2nd ed., John Wiley & Sons, New York, 1969.
52. *Handbook of Chemistry and Physics*, R. C. Weast, ed., CRC Press, Boca Raton, FL, 1986–1987, p. A-47, eq. 4.37.
53. P. Kraulis, MolScript, Version 1.1, *J. Appl. Cryst.*, **24**, 946 (1991).